

#3

Attorney Docket No. 1454.1067

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Horst-Udo HAIN

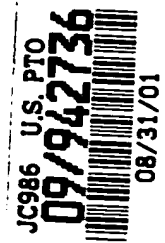
Application No.: NEW

Group Art Unit: To be assigned

Filed: August 31, 2001

Examiner: To be assigned

For: METHOD FOR SPEECH SYNTHESIS



**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN  
APPLICATION IN ACCORDANCE  
WITH THE REQUIREMENTS OF 37 C.F.R. §1.55**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. §1.55, the applicant(s) submit(s) herewith a certified copy of the following foreign application:

German Patent Application No. 100 42 942.4

Filed: August 31, 2000

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing date(s) as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. §119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: August 31, 2001

By: Mark J. Henry  
Mark J. Henry  
Registration No. 36,162

700 11th Street, N.W., Ste. 500  
Washington, D.C. 20001  
(202) 434-1500

**THIS PAGE BLANK (USPTO)**



## Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

**Aktenzeichen:** 100 42 942.4

**Anmeldetag:** 31. August 2000

**Anmelder/Inhaber:** Siemens Aktiengesellschaft, München/DE

**Bezeichnung:** Verfahren zur Sprachsynthese

**IPC:** G 10 L 13/00

**Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.**

München, den 30. März 2001  
**Deutsches Patent- und Markenamt**  
**Der Präsident**  
Im Auftrag

Ebert

**THIS PAGE BLANK (USPTO)**

## Beschreibung

## Verfahren zur Sprachsynthese

- 5 Die Erfindung betrifft ein Verfahren, eine Anordnung und ein Computerprogrammprodukt zur Sprachsynthese mittels Graphem-Phonem-Umsetzung.

Sprachverarbeitungsverfahren sind beispielsweise aus US  
10 6 029 135, US 5 732 388, DE 19636739 C1 und DE 19719381 C1 bekannt. In nicht gesprochener Form gespeicherter Text lässt sich über eine Sprachsynthese als Sprache ausgeben. Dazu werden in der Regel die einzelnen Wörter des Textes in einer Datenbank gesucht, die die phonetischen Transkriptionen zahl-  
15 reicher Wörter enthält. Die phonetischen Transkriptionen der in der Datenbank gefundenen Wörter werden zusammengesetzt und können als Sprache ausgegeben werden.

Da aber keine Datenbank vollständig ist, was in der Regel zur  
20 Reduktion der Datenbankgröße durchaus beabsichtigt ist, kommt es immer wieder vor, dass ein Text Wörter enthält, die in der Datenbank nicht gefunden werden. Diese Wörter werden dann mit einer Out-of-Vocabulary-Behandlung (OOV-Behandlung) phonetisch transkribiert. Dabei wird jedes Wort jeweils aus den  
25 einzelnen Buchstaben des Wortes zugeordneten Phonemen zusammengesetzt. Solche OOV-Behandlungen sind allerdings relativ rechenintensiv und führen in aller Regel zu schlechteren Ergebnissen als die phonetische Transkription ganzer Wörter aufgrund von Datenbankeinträgen.

30 Weiterhin ist es bekannt, die phonetische Transkription eines gegebenen Wortes aus den phonetischen Transkriptionen seiner Teilwörter zusammenzusetzen, wenn das gegebene Wort ausschließlich aus diesen Teilwörtern besteht.

35 Davon ausgehend liegt der Erfindung die Aufgabe zugrunde, Sprachsynthese dahingehend zu verbessern, dass in größerem

Umfang auf in einer Datenbank angegebene phonetische Transkriptionen von Wörtern zurückgegriffen werden kann und nur noch in geringerem Maße OOV-Behandlungen verwendet werden müssen.

5

Diese Aufgabe wird durch ein Verfahren, eine Anordnung und ein Computerprogrammprodukt mit den Merkmalen der unabhängigen Patentansprüche gelöst.

10 Durch das Verfahren, die Anordnung oder das Computerprogrammprodukt ist es möglich, auch dann für ein gegebenes Wort auf die phonetischen Transkriptionen seiner Teilwörter zurückzugreifen, wenn sich das gegebene Wort nicht vollständig aus in der Datenbank enthaltenen Teilwörtern zusammensetzen lässt.

15 Der wesentliche Gedanke ist dabei, dass erstmals eine hybride Vorgehensweise zum Einsatz kommt, bei der für dasselbe gegebene Wort sowohl die phonetische Transkription vollständiger Teilwörter, als auch eine OOV-Behandlung zum Einsatz kommt.

20 In einer bevorzugten Weiterbildung erfolgt die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils in Abhängigkeit der phonetischen Transkription des gefundenen Teilwortes. Hierdurch lässt sich die Qualität der Sprachsynthese für den weiteren Bestandteil gegenüber einer entsprechenden reinen OOV-Behandlung des gesamten Wortes deutlich

25 steigern. Dies liegt zum einen daran, dass die phonetische Transkription des gefundenen Teilwortes sehr viel sicherer ist als es eine phonetische Transkription dieses Teilwortes durch einen OOV-Behandlung wäre. Dadurch kann bei der OOV-

30 Behandlung des weiteren Bestandteils von einem sicheren phonetischen Kontext ausgegangen werden, was die OOV-Behandlung mit sehr viel größerer Wahrscheinlichkeit zum richtigen Ergebnis kommen lässt. Zum anderen ist die phonetische

35 Transkription des gefundenen Teilwortes sehr viel länger als die üblicherweise bei einer OOV-Behandlung verwendeten Phone-me. Deshalb ist der phonetische Kontext nicht nur sicherer, sondern auch länger, so dass die OOV-Behandlung für den wei-

teren Bestandteil aufgrund einer größeren Menge relevanter Informationen durchgeführt werden kann. Dieser Vorteil muss für die beanspruchte bevorzugte Weiterbildung allerdings nicht unbedingt ausgenutzt werden. Unter bestimmten Bedingungen kann es auch sinnvoll sein, wenn für die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils in Abhängigkeit der phonetischen Transkription des gefundenen Teilwortes lediglich der Teil des Teilwortes berücksichtigt wird, der dem weiteren Bestandteil unmittelbar benachbart ist.

Besonders vorteilhaft wird das Verfahren, wenn es nicht nach dem Auffinden eines ersten Teilwortes abgebrochen wird, sondern wenn noch weitere Teilwörter im gegebenen Wort gesucht werden. So wird ein möglichst großer Abschnitt des gegebenen Wortes aus Teilwörtern zusammengesetzt, für die in der Datenbank zuverlässige Informationen vorliegen, und lediglich der verbleibende, meist kleine weitere Bestandteil des Wortes muss einer OOV-Behandlung unterzogen werden.

Steht dieser verbleibende weitere Bestandteil zwischen zwei gefundenen Teilwörtern so wird die OOV-Behandlung vorzugsweise in Abhängigkeit beider gefundenen Teilbereiche vorgenommen. Dann ist nämlich sowohl der linke als auch der rechte phonetische Kontext des weiteren Bestandteils sicher vorgegeben, weshalb sich die OOV-Behandlung mit exzellenten Ergebnissen durchführen lässt.

Die Suche nach Teilwörtern in der Datenbank lässt sich durch verschiedene Maßnahmen optimieren. So sollte zum Beispiel nur nach Teilwörtern gesucht werden, die eine vorgegebene Mindestlänge aufweisen. Als Mindestlänge hat sich in der Praxis eine Länge von 5 Buchstaben herausgestellt, wobei bei anderen Rahmenbedingungen, zum Beispiel für eine andere Sprache, auch Mindestlängen von 3, 4 oder 6 Buchstaben sinnvoll sein können.

Weiterhin wird das Suchergebnis verbessert, wenn die Suche für einen Wortteil des gegebenen Wortes nicht sofort nach dem Auffinden des ersten passenden Teilwortes abgebrochen wird, sondern noch nach anderen möglichen Teilwörtern gesucht wird. Dies kann zum Beispiel erfolgen, indem der Wortteil um weitere Buchstaben ergänzt wird. In der Regel ergibt sich bei dieser Vorgehensweise das beste Ergebnis, wenn von mehreren gefundenen Teilwörtern dasjenige ausgewählt wird, das am längsten ist. Es kann allerdings auch ein kürzeres Teilwort ausgewählt werden, wenn dieses kürzere Teilwort in Verbindung mit einem in der Datenbank gefundenen und im gegebenen Wort enthaltenen längeren Teilwort einen größeren Teil des gegebenen Wortes darstellt, als das gefundene längere Teilwort für sich alleine, wenn dieses nicht mit dem gefundenen zweiten Teilwort kombiniert werden kann.

Die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils kann mittels eines neuronalen Netzes erfolgen.

Alternativ oder ergänzend kann für die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils eine regelbasiertes Verfahren oder ein DTW-Verfahren zum Einsatz kommen. Ein solches Verfahren ist zum Beispiel in Rüdiger Hoffmann „Signalanalyse und -erkennung“, Springer Verlag, Berlin, 1998, beschrieben.

Die OOV-Behandlung kann allerdings auch mittels einer zweiten Datenbank erfolgen, die die phonetischen Transkription von bei zusammengesetzten Wörtern üblicherweise verwendeten Füllpartikeln enthält. Im Deutschen sind dies insbesondere Dativ- und Genitivendungen, die bei zusammengesetzten Wörtern an das jeweils vorne stehende Wort angehängt werden.



Weitere wesentliche Merkmale und Vorteile der Erfindung ergeben sich aus der Beschreibung eines Ausführungsbeispiels anhand der Zeichnung; dabei zeigt

- 5    Figur 1 eine schematische Darstellung des Ablaufs des Verfahrens und

Figur 2 eine schematische Darstellung eines zwischen zwei  
Teilwörtern stehenden weiteren Bestandteils eines ge-  
gebenen Wortes.

Das Verfahren soll am Beispiel des gegebenen deutschen Wortes "Trainingslager" erläutert werden. Es sollen nur Teilwörter mit einer Mindestlänge von fünf Buchstaben gesucht werden. Im  
15    Schritt S1 gemäß Figur 1 wird für das gegebene Wort in einer Datenbank, die phonetische Transkriptionen von Wörtern enthält, nach Teilwörtern des gegebenen Wortes gesucht. Da die Mindestlänge auf fünf Buchstaben gesetzt ist, wird mit der Suche nach dem Wort "Train" angefangen. In einer deutschsprachigen Datenbank wird dieses Wort nicht gefunden. Enthält die  
20    Datenbank auch englischsprachige Wörter, so ist bereits jetzt das erste Teilwort des gegebenen Wortes gefunden. Vorzugsweise wird aber nicht nur im ersten, sondern auch im zweiten Fall weitergesucht. Dies geschieht durch die Suche nach dem  
25    Wort "Traini". Diese Buchstabenkombination wird in der Datenbank nicht gefunden. Das Gleiche gilt für die danach gesuchte Buchstabenkombination "Trainin".

Dagegen wird die nächste Buchstabenkombination "Training" in  
30    der Datenbank gefunden. Trotzdem wird auch in diesem Fall vorzugsweise weitergesucht, nämlich nach der Buchstabenkombination "Trainings" und den in entsprechender Fortsetzung dieses Suchschrittes gebildeten längeren Buchstabenkombinationen des gegebenen Wortes. Unter der Voraussetzung, dass das ge-  
35    bene Wort "Trainingslager" in seiner Gesamtheit nicht in der Datenbank gefunden wird, werden keine weiteren Teilwörter in der Datenbank gefunden.

Für den Fall einer englisch- und deutschsprachigen Datenbank wird aus den beiden gefundenen Teilwörtern "Train" und "Training" das längere Teilwort "Training" ausgewählt. Dieser Auswahlsschritt entfällt im Beispielfall für eine rein deutschsprachige Datenbank.

Im Schritt S3 wird für das gefundene Teilwort "Training" die in der Datenbank verzeichnete phonetische Transkription gewählt.

Gemäß dem Schritt S4 wird festgestellt, dass das gegebene Wort "Trainingslager" neben dem gefundenen Teilwort "Training" einen weiteren Bestandteil "slager" aufweist, der nicht in der Datenbank verzeichnet ist.

Dieser weitere Bestandteil "slager" wird dann im Schritt S5 mittels einer OOV-Behandlung phonetisch transkribiert. Diese OOV-Behandlung basiert vorzugsweise auf einer Umsetzung der einzelnen Grapheme des weiteren Bestandteils "slager" in Phoneme mittels eines neuronalen Netzes. Die Phoneme werden durch das neuronale Netz so ausgewählt und zusammengesetzt, dass sich für den weiteren Bestandteil für sich genommen eine möglichst gute Sprachsynthese ergibt.

Für ein noch besseres Sprachsyntheseergebnis erfolgt die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils "slager" in Abhängigkeit der aus der Datenbank gewählten phonetischen Transkription des gefundenen Teilwortes "Training". Das gefundene Teilwort "Training" bzw. seine phonetische Transkription gibt im gewählten Beispiel den linken phonetischen Kontext des weiteren Bestandteils "slager" sicher vor. Das für die OOV-Behandlung des weiteren Bestandteils "slager" verwendete neuronale Netz kann deshalb von einem sicheren Ergebnis der dem weiteren Bestandteil vorausgegangenen Silben des gegebenen Wortes ausgehen und ein ent-

sprechend sicheres Ergebnis für die phonetische Transkription des weiteren Bestandteils liefern.

5 Im letzten Schritt S6 des Verfahrens zur Sprachsynthese wird schließlich die phonetische Transkription des gefundenen Teilwortes "Training" und die phonetische Transkription des weiteren Bestandteils "slager" zusammengesetzt.

10 Das Sprachsyntheseergebnis lässt sich weiter verbessern, wenn nicht nur vom Anfang des gegebenen Wortes beginnend Teilwörter gesucht werden, sondern die Suche auch von anderen Bereichen des gegebenen Wortes aus gestartet wird. Ist für das Teilwort eine bestimmte Mindestlänge  $i$  vorgegeben, so empfiehlt sich der Beginn der weiteren Suche beim  $i$ -ersten Buchstaben. Im gegebenen Beispiel wird dann für  $i=5$  die weitere  
15 Suche mit der Buchstabenfolge "ingsl" gestartet, die ihrerseits wiederum die gegebene Mindestlänge aufweist. Diese Buchstabenfolge würde in der Datenbank nicht gefunden. Das gleiche gilt für die danach gesuchten Buchstabenfolgen "ings-  
20 la", "ingslag" usw.

Da bei dieser weiteren Suche keinerlei Teilwort gefunden wird, wird die darauffolgende Suche nicht beim Buchstaben  $2*i+1$  gestartet, sondern schon bei  $i+2$ . Allerdings führt auch  
25 die Suchsequenz "ngsla", "ngslag" usw. zu keinem Ergebnis. Nach dem Durchführen weiterer entsprechender Suchen wird allerdings in der letzten Suche das weitere Teilwort "lager" gefunden. Dieses weitere gefundene Teilwort "lager" entstammt nicht dem Wortteil des Wortes "Trainingslager", für den das  
30 erste Teilwort "Training" gefunden wurde. Deshalb muss im Beispiel keine Auswahl zwischen den beiden Teilwörtern getroffen werden.

Vielmehr verbleibt als weiterer Bestandteil des gegebenen  
35 Wortes "Trainingslager" nurmehr der Buchstabe "s". Dieser einzelne Buchstabe "s" lässt sich sehr leicht mittels einer OOV-Behandlung phonetischen transkribieren. Hierbei kommt

noch erleichternd hinzu, dass gemäß Figur 2 für das Zentrum 2 "s" sowohl der linke Kontext 1 "Training" als auch der rechte Kontext 3 "lager" bekannt ist.

- 5 Statt der OOV-Behandlung durch ein neuronales Netz, wie sie oben beschrieben wurde, kann in diesem Fall die OOV-Behandlung auch durch eine Suche in einer weiteren Datenbank erfolgen, in der die phonetischen Transkriptionen von bei zusammengesetzten Wörtern üblicherweise verwendeten Füllpartikeln enthalten sind. Das Genitiv-s des vorliegenden Beispiels  
10 ist ein solcher üblicherweise verwendeter Füllpartikel. Es würde deshalb in der zweiten Datenbank gefunden und die zugehörige phonetische Transkription gewählt werden.
- 15 Alternativ lassen sich für die OOV-Behandlung aber auch regelbasierte Verfahren und DTW-Verfahren einsetzen. In jedem Fall sind bessere phonetische Transkriptionen des weiteren Bestandteils zu erwarten, wenn bei der OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils die phonetische Transkription mehrerer oder aller gefundenen Teilwörter berücksichtigt wird. Dies ist natürlich insbesondere  
20 der Fall, wenn der weitere Bestandteil im Wort zwischen zwei gefundenen Teilwörtern angeordnet ist.
- 25 In einem letzten Schritt wird dann schließlich die phonetische Transkription des gefundenen Teilworts "Training", die phonetische Transkription des weiteren gefundenen Teilwortes "lager" und die phonetische Transkription des weiteren Bestandteils "s" zur Sprachsynthese zusammengesetzt.

30

Die erfindungsgemäße Anordnung lässt sich in Form eines Computersystems realisieren, das programmiert wird, ein entsprechendes Verfahren auszuführen.

## Patentansprüche

1. Verfahren zur Sprachsynthese mittels einer Graphem-Phonem-Umsetzung, bei dem

- 5 - für ein gegebenes Wort in einer Datenbank, die phonetische Transkriptionen von Wörtern enthält, nach Teilwörtern des gegebenen Wortes gesucht wird,
- mindestens ein Teilwort des gegebenen Wortes in der Datenbank gefunden wird,
- 10 - für das gefundene Teilwort eine in der Datenbank verzeichnete phonetische Transkription gewählt wird,
- das gegebene Wort neben dem gefundenen Teilwort mindestens einen weiteren Bestandteil aufweist, der nicht in der Datenbank verzeichnet ist,
- 15 - dieser weitere Bestandteil mittels einer OOV-Behandlung phonetisch transkribiert wird und
- die phonetische Transkription des gefundenen Teilwortes und die phonetische Transkription des weiteren Bestandteils zusammengesetzt wird.

20

2. Verfahren zur Sprachsynthese nach Anspruch 1, bei dem

- die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils in Abhängigkeit der phonetischen Transkription des gefundenen Teilwortes erfolgt.

25

3. Verfahren zur Sprachsynthese nach Anspruch 1 oder 2, bei dem

- für das gefundene Wort in der Datenbank nach weiteren Teilwörtern des Wortes gesucht wird,
- 30 - mindestens ein weiteres Teilwort des gegebenen Wortes in der Datenbank gefunden wird,
- für dieses gefundene weitere Teilwort eine in der Datenbank verzeichnete phonetische Transkription gewählt wird und
- die phonetische Transkription des gefundenen Teilworts, die
- 35 phonetische Transkription des weiteren gefundenen Teilworts und die phonetische Transkription des weiteren Bestandteils zusammengesetzt wird.

4. Verfahren zur Sprachsynthese nach Anspruch 3, bei dem  
- der weitere Bestandteil im gegebenen Wort zwischen dem gefundenen Teilwort und dem gefundenen weiteren Teilwort angeordnet ist und  
5 - die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils in Abhängigkeit der phonetischen Transkription des gefundenen Teilwortes und der phonetischen Transkription des gefundenen weiteren Teilworts erfolgt.
- 10 5. Verfahren zur Sprachsynthese nach zumindest einem der vorstehenden Ansprüche, bei dem  
- nur nach Teilwörtern gesucht wird, die mindestens eine vor-  
gegebene Mindestlänge aufweisen.
- 15 6. Verfahren zur Sprachsynthese nach zumindest einem der vorstehenden Ansprüche, bei dem  
- wenn für denselben Wortteil des gegebenen Wortteiles mehrere Teilwörter gefunden werden, von diesem das längste Teil-  
20 wort ausgewählt wird.
7. Verfahren zur Sprachsynthese nach zumindest einem der vorstehenden Ansprüche, bei dem  
- die OOV-Behandlung zur phonetischen Transkription des wei-  
25 teren Bestandteils mittels eines neuronalen Netzes erfolgt.
8. Verfahren zur Sprachsynthese nach zumindest einem der vorstehenden Ansprüche, bei dem  
- die OOV-Behandlung zur phonetischen Transkription des wei-  
30 teren Bestandteils mittels eines regelbasierten Verfahrens erfolgt.
9. Verfahren zur Sprachsynthese nach zumindest einem der vorstehenden Ansprüche, bei dem  
35 - die OOV-Behandlung zur phonetischen Transkription des weiteren Bestandteils mittels einer zweiten Datenbank erfolgt,

die die phonetischen Transkription von bei zusammengesetzten Wörtern üblicherweise verwendeten Füllpartikeln enthält.

10. Anordnung zur Sprachsynthese mittels einer Graphem-  
5 Phonem-Umsetzung, die so ausgebildet ist,  
- dass für ein gegebenes Wort in einer Datenbank, die phonetische Transkriptionen von Wörtern enthält, Teilwörter des gegebenen Wortes suchbar sind,  
- mindestens ein Teilwort des gegebenen Wortes in der Datenbank auffindbar ist,  
10 - dass für das gefundene Teilwort eine in der Datenbank verzeichnete phonetische Transkription auswählbar ist,  
- wobei das gegebene Wort neben dem gefundenen Teilwort mindestens einen weiteren Bestandteil aufweist, der nicht in der  
15 Datenbank verzeichnet ist,  
- dass dieser weitere Bestandteil mittels einer OOV-Behandlung phonetisch transkribierbar ist und  
- die phonetische Transkription des gefundenen Teilwortes und die phonetische Transkription des weiteren Bestandteils zusammensetzbar ist.  
20

11. Computerprogrammprodukt zur Sprachsynthese mittels einer Graphem-Phonem-Umsetzung, bei dem beim Ablauf auf zumindest eineressoreinheit  
25 - für ein gegebenes Wort in einer Datenbank, die phonetische Transkriptionen von Wörtern enthält, nach Teilwörtern des gegebenen Wortes gesucht wird,  
- mindestens ein Teilwort des gegebenen Wortes in der Datenbank gefunden wird,  
30 - für das gefundene Teilwort eine in der Datenbank verzeichnete phonetische Transkription gewählt wird,  
- das gegebene Wort neben dem gefundenen Teilwort mindestens einen weiteren Bestandteil aufweist, der nicht in der Datenbank verzeichnet ist,  
35 - dieser weitere Bestandteil mittels einer OOV-Behandlung phonetisch transkribiert wird und

- die phonetische Transkription des gefundenen Teilwortes und die phonetische Transkription des weiteren Bestandteils zusammengesetzt wird.



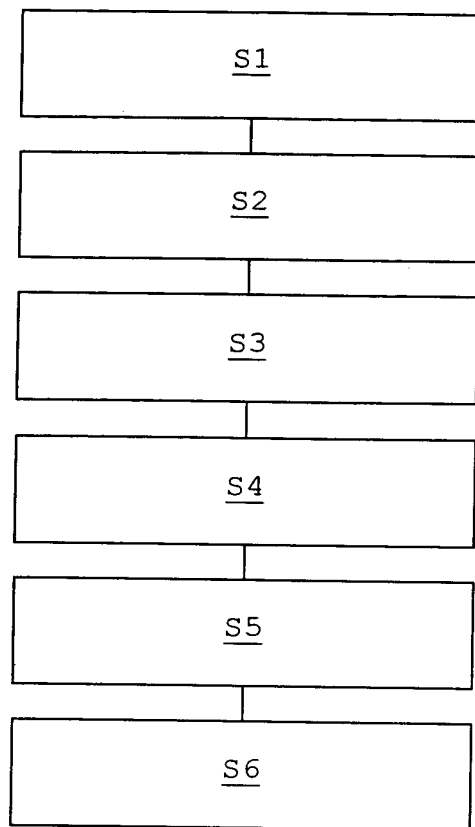
## Zusammenfassung

### Verfahren zur Sprachsynthese

- 5 Die Erfindung betrifft ein Verfahren, eine Anordnung und ein Computerprogrammprodukt zur Sprachsynthese mittels einer Graphem-Phonem-Umsetzung. Dabei wird für ein gegebenes Wort in einer Datenbank, die phonetische Transkriptionen von Wörtern enthält, nach Teilwörtern des gegebenen Wortes gesucht. Wird  
10 mindestens ein Teilwort des gegebenen Wortes in der Datenbank gefunden, so wird für das gefundene Teilwort eine in der Datenbank verzeichnete phonetische Transkription gewählt. Das gegebene Wort weist neben dem gefundenen Teilwort mindestens einen weiteren Bestandteil auf, der nicht in der Datenbank  
15 verzeichnet ist. Dieser weitere Bestandteil wird mittels einer OOV-Behandlung phonetisch transkribiert und die phonetische Transkription des gefundenen Teilwortes und die phonetische Transkription des weiteren Bestandteils wird zusammengesetzt.

20

Figur 1



Figur 2

